Environmental Researchers' Data Practices: An Exploratory Study in Turkey*

Suzie Allard[1] and Arsev Umur Aydınoğlu[2]

[1] University of Tennessee, School of Information Sciences, Knoxville, TN – 37996 USA
sallard@utk.edu
[2] NASA Astrobiology Institute, Moffett Field, CA – 94035 USA
arsevu@gmail.com

Abstract. This qualitative exploratory study probes the knowledge and attitudes of information science and environmental researchers in Turkey towards scientific data and information particularly in regards to sharing and preservation. Ten environmental scientists and two information scientists were interviewed. Results reveal that research data is stored mostly on personal computers. This raises two issues: organizational and technological sustainability. Environmental scientists in Turkey do not engage in data-intensive research. Data sharing is limited because of socio-cultural reasons. Although data collaboration is limited, there is an interest in international collaboration. According to the information scientists in Turkey, conversation on data management and data sharing has started but only in academic circles. Data practices are not mandated by funding agencies. The skill sets of information scientists are not fully utilized. For both environmental and information scientists, there is a long way to go.

Keywords: Environmental scientists in Turkey, data sharing, data preservation.

1 Introduction

Climate change is a grand challenge for science since the environmental impact touches societies across the globe [2] especially as society's consumption rates increase, populations grow, and nations modernize [3]. Climate change and its results have been identified by scientists [4], [5], [6] and shared with the public through the popular media [7], [8], [9], [10]. Climate change is implicated in rising sea levels [11] and erratic weather patterns [12] which can precipitate starvation and disease [13] , [14] , [15].

Challenges such as climate change and studies focusing on these challenges such as biodiversity require new approaches to science [16]. Scientific research is increasingly becoming more complex [17], including data-intensive science, which gains new insights through data-driven approaches [18]. Data-driven science includes

* This study was supported by a 2010 award from the College of Communication and Information (CCI) Dean's Summer Research Grant Program. An earlier version of this study was presented as a poster [1].

S. Kurbanoğlu et al. (Eds.): IMCW 2012, CCIS 317, pp. 13–24, 2012.

using data gathered from global locations and often uses computational modeling techniques to create new ways to understand the problems. The phrases "the fourth paradigm" or "data-intensive scientific discovery era" [19] have been quickly accepted among researchers as the science literature and data become more accessible online and are increasingly able to interoperate with each other [20]. Data has become more important as the advances in digital computing, remote sensing technologies, and storage technologies allow scientists to engage with the data in new ways to create new knowledge. The new technologies have increased the amount of data collected, used, re-used, and stored [21].

The benefits of data sharing and data reuse, and the importance of data preservation are well documented, The benefits of data sharing include: (i) verifying results, since the re-analysis of data is necessary to replicate studies which can be used in training of new researchers; (ii) reducing re-collection costs; (iii) increasing data integrity through preservation; (iv) reducing data availability reduces the risk of data falsification and fabrication; (v) facilitating new insights and understandings through integrating different datasets [22], [23], [24], [25], [26].

Problems can best be addressed if barriers – disciplinary and geographic – can be overcome [27], [28], [29]. Overcoming these barriers is especially important when studying grand challenges such as global climate change since these are complex systems [30] that require collaborative and interdisciplinary scientific approaches [31]. One approach is the emergence of virtual organizations, i.e. DataONE, that allow scientists to more easily communicate and share their data [32], [33], [34]. This means that we must understand more than the technical issues; we must also understand the socio-cultural, economic, ethical, and political issues that influence scientists' data practices on the global stage. To do this, it is helpful to understand how environmental scientists in different communities conduct their work– whether community is defined by scientific domain or geographic area.

This study focuses on environmental scientists in Turkey because Turkey is an important environmental region and because there is no literature reporting on these scientists' knowledge and attitudes towards data practices. Due to its geographic location and diverse climate, Turkey is rich with endemic species. Moreover, it is an important hub for migrating birds. Therefore, environmental data from Turkey is of importance not only to Turkish audiences but also European, Asian, Middle-Eastern, and African audiences. It is important to know data practices of Turkish researchers since they are collecting data that could inform environmental researchers in each of these other regions. In order to have a better understanding of the practices, we conducted qualitative research in Turkey which explores the knowledge and attitudes of information science and environmental researchers in Turkey towards scientific data and information, particularly in regards to sharing and preservation.

2 Methods

We took a grounded theory approach and used long interviews to allow the themes to emerge from the discussions of our participants [35]. We interviewed two groups of

participants in Turkish universities and research institutions: (1) environmental scientists; and (2) scholars of library and information sciences who are interested in database management, dissemination of information, information architecture, and knowledge management. The importance of environmental research is summarized above; thus, we chose to study this community. The library and information science community was selected because they could play the support role to create the necessary cyberinfrastructure for environmental scientists to take care of their research data. Two interviews were completed with library and information scientists and ten interviews were completed with environmental scientists. We were not able to find more interviewees from library and information science, hence the difference in numbers of interviews.

In the summer of 2010, we conducted in-depth interviews with participants in order to understand what library and information science (LIS) scholars and environmental scientists think about scientists and their data practices in Turkey. Personal contacts were used to initiate contact and snowball sampling was also used to recruit participants. One investigator had worked with NGOs in Turkey from 1999 to 2007 and established a personal network of environmental scientists. Potential participants were contacted by phone and asked if they were interested in participating in such a study. If they were, a meeting was scheduled, and an informed consent form was presented at the meeting prior to conducting the interview.

The interviews were conducted as informal conversations, which were guided by two discussion guides (one for each community) with several open-ended questions. Environmental scientists were asked whether and how they take care of their research data, and their perceptions towards data sharing and preservation. In addition to their research data habits, library and information scientists were asked what they think of scientists' research data practices and what the library and information science community can do to support scientists.

Of the twelve interviews we conducted, two of them were in English and ten were in Turkish. They were translated to English by the researcher. We used the English transcriptions in the analysis. There were separate discussion guides for the information science scholars and the environmental scientists. Some questions emerged from the initial interviews and were used in subsequent interviews.

The first few questions were designed to make the respondents feel more at ease with the interviewer and more comfortable in discussion. The subsequent questions asked the respondents to express their thoughts and feelings toward scientific data and information sharing and preservation. Interviews lasted between 30 to and 75 minutes. Each interview was audio-recorded and verbatim transcribed by the Co-Principal Investigator (Co-PI) for analyzing the data and quotes. The interviews were conducted at the researchers' offices (except one that was in a coffee shop) to ensure high quality audio recording and confidentiality. The audio recordings were destroyed after each interview was transcribed.

Analytic induction was used to find common patterns in the interviews by reviewing the transcripts line by line for themes or categories emerging from the initial cases, then modifying and refining them on the basis of subsequent cases. A peer audit technique, a technique in which researchers check and compare each other's analyses, was used to help clarify the analysis.

3 Results: Emerging Themes

3.1 Environmental Scientists

The ten interviews with environmental scientists exhibited redundancy and therefore provided rich results. There were six emerging themes:

3.2 About Data

Data is most important for publication. Most scientists mentioned that they were "done" with the data after their publication. This meant that at the time of data collection the scientist did not plan how to maintain the data after the analysis was completed and the results published. Although scientists we interviewed had some interest in data as a "research product", none was actively engaged in a formal process to preserve this product. Instead the process for maintaining the data was informal and was usually focused on storage activities conducted after publication rather than being part of a process begun at the inception of data collection. One example illustrating this is noted below,

> "Q. What happens to your data and findings after your research?
> R. We write papers, publish.
> Q. What else? What happens to your raw data?
> R. We store it. Now we are making a database.
> Q. Who are 'we'?
> R. Me and my students I'm talking about it with." – Respondent 2.2

3.3 Storage

Data storage activities exist but do not address sustainability. Most researchers keep their data on personal computers and related media such as CDs and external hard disks. As reflected in the comments below some researchers have an active backup strategy. However, these strategies are usually limited to keeping copies within the research group and often in one physical location. Additionally, these stored copies are only for the PI and the research group. Therefore when the PI retires, the data could be lost forever.

> "In office computers, hard disks, and CDs. The photos and others. We copy everything into CDs and hardcopy" – Respondent 2.3
> "They are in external hard disks and also in CDs. I have 2-3 copies in case something happens. I store such information in multiple copies." - Respondent 2.1
> "I don't use a server but keep flash disks. Not specifically for that data but, with panic, in case something happens. Sometimes I store it in a big external hard disk, sometimes I store it in pieces in small storage media. I try to backup data and my personal files." – Respondent 2.4

There are also substantial issues regarding technological sustainability. These issues often result from software becoming obsolete and researchers having no plan for data migration to new formats. Researchers note that often the data collected in previous years is not accessible anymore.

> "I even had a database about my field notebook made but now I can't use that software because the operating system has changed. … [T]he digital environment changes in 3 years. You have to stop all your work and try to keep up with the new stuff [format] because everyday something new is introduced. You have to change accordingly; thus, you have to deal with that only. Thus, my data that I put into my computer with the operation system I had in 91-92 is now unavailable/unusable." – Respondent 2.6
> "Yes, we have format problems with very old data. Even, for some time – I'm not a very young person so let me explain to you like this. There were some operating systems different from IBM PC. We have some simulation work done in these systems. There is nothing to make them work anymore. The floppy disks are here but we can't use them." – Respondent 2.9

3.4 Data-Intensive Research

Data-intensive science is not yet a regular part of the research environment. Turkish environmental scientists are not yet regularly engaging in data-intensive science for a variety of reasons.

> "What I'm trying to do is not interpreting something that is already known. Discovering a phenomenon, that's what I like. But, of course, the big datasets you mention might point out a phenomenon. I'm not denying that, I'm aware of that." –Respondent 2.7

But some are encouraging students to use data in new ways. In some cases, individual graduate students contact the PIs and ask for datasets. However, there is not an established mechanism to attract the interested researchers. It generally works as mouth of word.

> "What happens is a student comes in who is very keen on modeling and data mining, you know. Data can be used from different angles, different perspectives." –Respondent 2.2

3.5 Data Sharing

Data sharing is only engaged in on a limited basis. Data sharing happens primarily when multiple researchers are involved in a specific project. However, there is little sharing outside the research group which is involved in the initial data collection. The interviews suggest the reason extends beyond technological limitations to several socio-cultural reasons. Trust is one of the most prominent reasons among them.

> "In fact I am a sharing person, however, the trustworthiness of the person that I'm going to share with is critical. For instance, I might share with the people I work with or I trust. But in our country, besides plagiarism, there are cases that someone else's data is used and published. Thus, I am not sharing often." –Respondent 2.5

Additionally data sharing is limited by concerns for intellectual attribution.

> "I'd very much like to publish; however, if person A calls me and says 'you did such a study but you didn't publish. I'm going to do something like this, I'll benefit from them (the data)', I might not give it." – Respondent 2.4

The scientists we spoke with also feel that they do not receive enough support from their organizations. Limitations on resources, technology, staff, and the training needed to exercise stewardship of data means that the lack of institutional support becomes an important barrier.

> "You don't have continuation. Instead of this I'd rather work in an institute where you work with technicians and so on, supporting groups. You'd have a computer guy, database, graphic institute-office everything, whereas in a department like this [academic department], you do research with your own capacity, own students. You get nothing else as support." Respondent 2.2
> "Q. Does the university you work for provide you space on a server?"
> "No, no, no. No such thing." Respondent 2.5

In addition to limited resources, academic institutions are not promoting a culture of data sharing among the scientists. The scientists do not get any credit for the data they provide to others. In fact, many feel that their efforts and hard work are undermined.

> "At the universities, there isn't any institutional culture that encourages sharing. There, it depends on the skills of the individual only." – Respondent 2.6
> "The institution is of course not aware of such a thing. We do it ourselves, by our own methods. I mean, I don't know how the institution could provide support." – Respondent 2.8

In fact there was a general feeling that in a big dataset, their scientific contribution seems 'minor'; however, it does not feel 'minor' to them because of the time and effort they had to put into the data collection. This kind of inequity makes the scientists reluctant to share.

3.6 Data Collaboration

Collaboration with the broader scientific community is limited. Scientists felt that there is limited science data collaboration with researchers outside Turkey. However, the scientists expressed an interest in being involved in international collaborations.

> "I haven't had such an experience but I believe it should be happening. I think very positively about it. I believe the real sound studies could emerge in that way. Quite the opposite, I consider myself an introvert and unproductive because of this (not having international experience)." – Respondent 2.1
> "Many things in environment are transboundary. Thus, collaboration is a must; regional and international collaboration is a must." –Respondent 2.3

3.7 Academic vs. Government

Attitudes towards the use and storage of data vary with the research environment. Scientists in both the academic and government research environments mentioned incentives. The incentives seem to be one of the factors that influence scientists' attitudes towards data and data sharing.

> "In government, it's much easier to share information. You don't have many problems because ultimately you are an environment specialists, even if you write five articles you will continue to be an environment specialist, … Neither your salary will change, nor your title." – Respondent 2.4

3.8 Scholars of Information Science

Some of the barriers hindering the sustainable data practices and data sharing of Turkish environmental scientists might be overcome through collaborating with information specialists. Therefore, we interviewed information science scholars about their relationship with environmental scientists in Turkey and their impressions of the scientists' data practices. Most important is the information science scholars' attitude toward collaborating with the environmental scientists.

Although we contacted seven information science scholars, very few felt they could discuss data practices, particularly in reference to science information, so only two interviews could be conducted. The fact that so few felt qualified to discuss this topic suggests that there is potential for expanding the understanding of science data through increased training opportunities of information scholars. For the purposes of this study, the small number of interviews meant that redundancy could not be reached and themes could not be identified. However, the results are still valuable in providing a very preliminary look at the information science perspective and in providing details about the science data environment.

The information science scholars we interviewed believed that most scientists kept their data on personal computers. They also believed that data is seen as a means to reach an end—a publication; rather than as a research product that has value in and of itself. These beliefs aligned well with our findings from interviewing the environmental scientists.

> "I think information scientists should have dual roles in data preservation. First of all they … themselves [need] to practice you know data preservation. But in general information scientists are not that different from other scientists when it comes to data storage, data preservation, etc. because as a researcher they just want to complete this study and walk out just like any other scientist." – Respondent 1.1

According to the information science scholars, a conversation regarding data management and sharing has been simmering within some academic circles, but it is still only on a limited basis. While there is some interest in sharing there are concerns about who will implement and maintain the process for sharing because most scientists are not willing to do so. Additionally there are concerns about how data will be used.

Data practices that are mandated by a granting agency, for example a European Union collaborative research project, are likely to be adopted for that project but this does not mean these practices will be adopted for other work conducted by that scientist or research group. The information science scholars feel that information sciences can play an important role in supporting interdisciplinary science efforts. However, they feel there are hurdles to overcome in order for information science to play a vital role. They note this would require more people to be aware of information science and what it can do to support research.

> "I'm speaking for Turkey, I don't think our field is known by many." – Respondent 1.2

These scholars feel that as the scientific world moves towards more interdisciplinary and data intensive research, Turkey must become involved. This would be a new direction in Turkish library and information science education.

> "These are new topics for us, new collaboration topics. I think this is going to change the curriculum of the information sciences." –Respondent 1.1

4 Conclusion

This is the first study that deals with the data practices of environmental researchers in Turkey. Clearly, more studies are needed to make generalizations and implement policies accordingly. However, the findings resemble those found in previous studies done in different parts of world [36], [26]. Data practices are poor among many scientists and Turkish environmental scientists are not very different. Strong data practices include activities in all eight steps of the data lifecycle (Fig. 1). These steps promote the sustainability and accessibility of data into the future. Scientists engage in some of these steps and other steps are supported by information specialists.

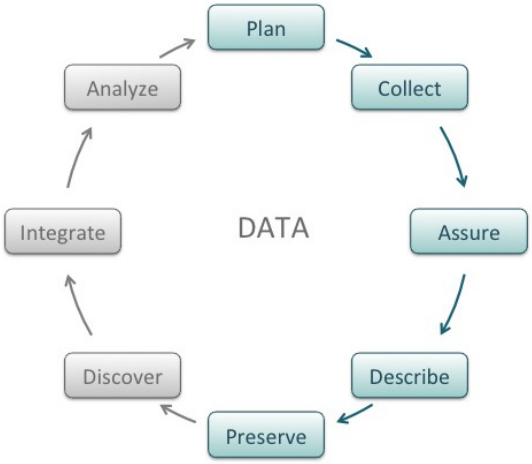

Fig. 1. Data life cycle [37]

The DataONE data lifecycle has eight unique stages (Fig. 1). It begins with creating the research plan, then progresses through data collection, quality assurance and quality control. Metadata is created to describe the data. The data is then deposited in a trusted repository where preservation activities may occur. Data discovery, integration, and analysis including visualization can then be supported by tools and services [37].

The results of this study suggest that, while Turkish environmental researchers and information science scholars participate in some steps of the data lifecycle, the existing research environment does not provide the organizational or technological infrastructure to support the full life cycle. For example, while several scientists have an active back-up strategy, they are conservative in promoting and sharing their data. Trust and intellectual attribution are the most mentioned socio-cultural barriers to data sharing. Moreover, lack of institutional support is an issue. Thus, they do not engage in data-intensive research. Metadata/interoperability issues and interdisciplinary barriers, which came up in previous studies, did not come up as data intensive science has not developed enough in Turkey.

In our opinion, research institutions and funding agencies could address some of the socio-cultural issues by providing incentives (such as having citation of a dataset increase points towards promotion) and resources (such as technology and training) for researchers and the information specialists who could help provide support for their intellectual pursuits. International collaboration opportunities should be utilized more since researchers are open to collaboration and there are some readily available EU funds for such opportunities. Promoting collaboration with information science scholars is a must to reach this goal; however, the information science community must do more to increase awareness among scientists of what information science offers and how that can positively impact the quality of the environmental scientists' research.

While this study provides insights into the data practices of environmental scientists in Turkey there is still much work that needs to be done to better understand the full picture. Our future research plans include preparing a survey based on the findings from these interviews which would measure the attitudes of environmental scientists in Turkey towards data sharing by accessing a much larger sample. These results could help the science policy makers in Turkey to develop relevant policies/incentives and could also help identify potential collaborators outside Turkey, which would help facilitate the collaboration process.

References

1. Allard, S., Aydinoglu, A.: Knowledge and Attitudes towards Scientific Data Practices & Preservation among Turkish Environmental Scientists and Information Scientists. Poster Presented at the 33rd Annual Research Symposium College of Communication and Information. University of Tennessee (2011), http://trace.tennessee.edu/cgi/viewcontent.cgi?article=1007&context=ccisymposium
2. United Nations - World Commission on Environment and Development. Our Common Future. UN Documents, New York (1987)
3. Bongaarts, J.: Population Growth and Global Warming. Population and Development Review 18, 299–319 (1992)
4. Intergovernmental Panel on Climate Change (IPCC): IPCC Fourth Assessment Report: Climate Change 2007 (2007), http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml#1
5. Krauss, W., Van Storch, H.: Culture Contributes to Perceptions of Climate Change. Nieman Reports 59, 99–102 (2005)
6. Victor, D.: The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming. Princeton University Press, Princeton (2004)
7. Egenter, S.: Sea Levels Could Rise More Than a Meter by 2100 (2009), http://www.Reuters.com
8. Morrello, L.: Study Finds Big Storms on a 1,000-Year Rise. New York Times (2009)
9. Revkin, A.C., Broder, J.M. In: Face of Skeptics, Experts Affirm Climate Peril. The New York Times (2009), http://www.nytimes.com/2009/12/07/science/earth/07climate.html?_r=2&ref=todayspaper
10. Russell, C.: Climate Change: Now What? Columbia Journalism Review 47, 45–49 (2008)
11. Meehl, G., Washington, W.M., Arblaster, J.M., Hu, A., Buja, L.E., et al.: How Much More Global Warming and Sea Level Rise? Science 307, 1769–1772 (2005)
12. Shah, A.: Climate Change and Global Warming. Global Issues.org (2009)
13. Gopalakrishnan, R.: Climate Change a Factor in Deaths from Disease: WHO (2009), http://www.Reuters.com
14. Nobel Laureates. The Next Hundred Years. Journal of Public Health Policy 23, 8 (2002)
15. Patz, J.A., Epstein, P.R., Burke, T.A., Balbus, J.M.: Global Climate Change and Emerging Infectious Diseases. Journal of the American Medical Association 275, 217–223 (1997)
16. Kelling, S., Hochachka, W., Fink, D., Riedewald, M., Caruana, R., Ballard, G., et al.: Data-intensive Science: A New Paradigm for Biodiversity Studies. BioScience 59, 613–620 (2009)
17. Lynch, C.: Big Data: How do Your Data Grow? Nature 455, 28–29 (2008)

18. Newman, H.B., Ellisman, M.H., Orcutt, J.A.: Data-intensive e-science Frontier Research. Communications of the ACM 46, 68–77 (2003)
19. Gray, J.: Jim Gray on eScience: A Transformed Scientific Method. In: Hey, T., Tansley, S., Tolle, K. (eds.) The Fourth Paradigm: Data-intensive Scientific Discovery (2009), http://tamingdata.com/2009/12/16/the-fourth-paradigm-data-intensive-scientific-discovery/
20. Hey, T., Tansley, S., Tole, K. (eds.): The Fourth Paradigm: Data-intensive Scientific Discovery (2009), http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf
21. National Academies of Science, Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age (2009), http://www.nap.edu/catalog.php?record_id=12615
22. Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., et al.: Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Data Science Journal 3, 135–153 (2004)
23. European Science Foundation: Shared Responsibilities in Sharing Research Data: Policies and Partnerships (2007), http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/Science_Policy/Sharing_Data/ESF_DOC_SHARINGDATA_V01ppp.pdf&t=1287487773&hash=89d4ad56d3544ba3edb68c1683369ed4
24. Inter-University Consortium for Political and Social Research. Guide to Social Science Data Preparation and Archiving: Best Practice throughout the Data Life Cycle. University of Michigan, Ann Arbor (2009), http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf
25. National Science Foundation, Office of Cyberinfrastructure Directorate for Computer & Information Science & Engineering. Sustainable Digital Data Preservation and Access Network Partners (DataNet) Program Solicitation - NSF 07-601 (2008), http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141 (retrieved September 5, 2010)
26. PARSE Insight: PARSE Insight (2009), http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
27. Allard, S.: Erasing the Barrier Between Minds: Freeing Information, Integrating Knowledge. American Communication Journal 4 (2001)
28. Borgman, C.L. (ed.): Scholarly Communications and Bibliometrics. Sage Publications, Newbury Park (1990)
29. Geertz, C.: Local Knowledge: Further Essays in Interpretive Anthropology. Basic Books, Inc., Publishers, New York (1983)
30. Rind, D.: Complexity and Climate. Science 284, 105–107 (1999)
31. Allard, S., Allard, G.: Transdisciplinarity and Information Science in Earth and Environmental Science Research. In: Proceedings of the 90th Annual Meeting of the American Society of Information Science and Technology, Vancouver, B.C., Canada, November 6-11 (2010), http://www.asis.org/Conferences/AM09/posters/46.doc
32. Allard, S., Tenopir, C., Wilson, B.: DataNetONE (Observation Network for Earth): An Earth Environmental and Ecological Sciences Data Center from a Communication and Information Perspective. In: Proceedings of Thirty-First Annual Communications Research Symposium. University of Tennessee, Knoxville (2009)

33. Michener, W.: Building Informatics Solutions for Multi-Decadal Ecological Research: Re-envisioning Science, Technology, and the Academic Culture. Presentation at the National Center for Ecological Analysis and Synthesis (2009)
34. Michener, W., Allard, S., Cobb, J., Cook, R., Cruse, P., Frame, M., et al.: DataNetONE: Enhancing Data-intensive Biological and Environmental Research through Cyberinfrastructure. Poster Presented at NASA Terrestrial Ecology Meeting, La Jolla, CA, March 15-17 (2010)
35. Strauss, A., Corbin, J.: Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, 2nd edn. Sage Publications, Thousand Oaks (1998)
36. Tenopir, C., Allard S., Douglass K., Aydinoglu, A.U., Wu, L., Read, E., et al.: Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6 (2011)
37. DataONE (2011), https://www.dataone.org